

STEFANO BERETTA

ste.beretta@gmail.com

Milano, Italia

TITOLI DI STUDIO

Dottorato di Ricerca

22/02/2013

Informatica

Milano, Italia

- Dottorato di ricerca in informatica (XXV ciclo) conseguito presso l'Università degli Studi di Milano-Bicocca.
- Il titolo della tesi è “*Algorithms for Next Generation Sequencing Data Analysis*”, svolta sotto la supervisione di Prof.ssa Paola Bonizzoni e Prof. Gianluca Della Vedova.

Laurea Specialistica

28/04/2009

Informatica

Milano, Italia

- Conseguita presso l'Università degli Studi di Milano-Bicocca, con votazione di 110/110 e Lode.
- Il titolo della tesi è “*Algoritmi per la Predizione della Struttura Secondaria dell'RNA*”, svolta sotto la supervisione di Prof.ssa Paola Bonizzoni e Prof. Gianluca Della Vedova.

Laurea Triennale

16/10/2006

Informatica

Milano, Italia

- Conseguita presso l'Università degli Studi di Milano-Bicocca, con votazione di 107/110.
- Il titolo della relazione finale è “*Un framework per la simulazione di sistemi complessi: metodologie di visualizzazione scientifica*”, svolta sotto la supervisione di Prof. Gianpiero Cattaneo e Dott. Fabio Mario Farina.

COLLABORAZIONI

Partecipazione a Progetti di Ricerca

- PRIN 2010-2011 dal titolo “*Automi e Linguaggi Formali: Aspetti Matematici e Applicativi*”, coordinatore scientifico: Prof. Antonio Restivo.
- Unità: Università degli Studi di Milano-Bicocca, responsabile scientifico: Prof.ssa Paola Bonizzoni.

Borse di Studio

- Borsa di Studio Ministeriale per la frequenza del Dottorato di Ricerca da Gennaio 2010 a Dicembre 2012.

ATTIVITÀ SCIENTIFICA

Descrizione dell'attività di ricerca

- La mia attività di ricerca si colloca in ambito bioninformatico. I miei principali interessi riguardano il disegno, lo sviluppo e la sperimentazione di algoritmi combinatori per la soluzione di problemi in genomica e trascrittomica. Nello specifico, lavoro alla realizzazione di applicazioni per l'analisi di sequenza biologiche (DNA ed RNA).

Durante il mio dottorato di ricerca mi sono occupato dello sviluppo di un modello computazionale per la caratterizzazione di eventi di splicing alternativo, a partire da dati di sequenziamento di nuova generazione (RNA-Seq). Obiettivo di tale ricerca è quello di ricostruire a partire da dati RNA-Seq, in presenza o assenza di un genoma di riferimento, la struttura dei geni espressi. Tale struttura è

finalizzata a rappresentare la composizione di trascritti dei geni, che sono soggetti ai meccanismi di splicing alternativo.

Ho inoltre svolto ricerca su pipeline per l'analisi di dati di sequenziamento trascrittomici, combinando metodi per l'allineamento di sequenze come TopHat e GSNAP, a tool per la ricostruzione di trascritti, come ad esempio Cufflinks e SLIDE. Tali pipeline sono state valutate mediante una comparazione diretta con dati simulati e reali, provenienti dal progetto ENCODE, sulla base dei risultati ottenuti.

Attualmente la ricerca è orientata al perfezionamento del tool sviluppato nell'attività di tesi ed in futuro si vuole approfondire la ricostruzione di trascritti facendo riferimento ad annotazioni di cDNA o basandosi su informazioni provenienti da altri genomi di riferimento.

Durante l'attività di ricerca svolta ho acquisito competenze relative all'utilizzo di banche dati genomiche, come ad esempio NCBI, sia in termini di interrogazione e navigazione finalizzata al reperimento dei dati che di manipolazione di questi ultimi.

Sono anche interessato all'integrazione di tool di predizione della struttura genica con tool di predizione di eventi di variazione strutturale, con l'obiettivo di integrare i due approcci al fine di migliorare la qualità dei risultati prodotti.

Infine, durante il mio periodo presso la *Technical University of Catalonia* (UPC) sotto la supervisione di Prof. Gabriel Valiente, mi sono occupato di studi relativi a problemi di metagenomica. In particolare ho lavorato allo sviluppo di algoritmi per il corretto assegnamento di sequenze genomiche, provenienti da organismi presenti in un campione analizzato, in alberi tassonomici. Questo lavoro di carattere sia algoritmico che applicativo, mi ha permesso di sviluppare delle competenze in ambito metagenomico riguardanti tecniche di assegnamento di sequenze e comparazione di alberi tassonomici. A questo proposito, mi sono occupato dell'analisi delle principali banche dati metagenomiche, quali NCBI Taxonomy, RDP, Greengenes e SILVA.

COMPETENZE INFORMATICHE

Durante il mio dottorato di ricerca in informatica ho sviluppato competenze informatiche in diversi linguaggi programmatici e di scripting, quali C/C++, Java, Perl e Python. Ho inoltre acquisito conoscenze dei linguaggi R e Gnuplot per l'analisi statistica e del linguaggio TeX/LaTeX per la realizzazione di documenti testuali.

Ho infine competenze nell'utilizzo dei sistemi operativi Unix e Windows e della suite Microsoft Office.

ELENCO DELLE PUBBLICAZIONI

Articoli su Riviste Internazionali

- [1] Daniel Alonso-Aleman, Aurélien Barré, **Beretta, Stefano**, Paola Bonizzoni, Macha Nikolski, and Gabriel Valiente. "Further Steps in TANGO: Improved Taxonomic Assignment in Metagenomics". *Bioinformatics* (2013). DOI: 10.1093/bioinformatics/btt256.

La copia della pubblicazione dal titolo: "Further Steps in TANGO: Improved Taxonomic Assignment in Metagenomics" riprodotta per intero come versione accettata per la pubblicazione in *Bioinformatics* composta di n. 7 pagine è conforme all'originale.

Atti di Conferenze Internazionali

- [2] **Beretta, Stefano**, Paola Bonizzoni, Raffaella Rizzi, and Gianluca Della Vedova. "Reconstructing isoform graphs from RNA-Seq data". In: *Bioinformatics and Biomedicine (BIBM), 2012 IEEE International Conference on*. 2012, 1–4. DOI: 10.1109/BIBM.2012.6392734.

La copia della pubblicazione dal titolo: “Reconstructing isoform graphs from RNA-Seq data” riprodotta per intero come versione accettata per la pubblicazione in *2012 IEEE International Conference on Bioinformatics and Biomedicine* composta di n. 4 pagine è conforme all’originale.

Workshop Internazionali (con peer-review)

- [3] Anna Paola Carrieri, **Beretta, Stefano**, Gianluca Della Vedova, Ernesto Picardi, Yuri Pirola, Raffaella Rizzi, Graziano Pesole, and Paola Bonizzoni. “An in-silico framework for comparing and validating transcripts predicted from single and paired-end reads”. In: *Next Generation Sequencing Workshop*. 2012.
- [4] Yuri Pirola, Raffaella Rizzi, **Beretta, Stefano**, Ernesto Picardi, Graziano Pesole, Gianluca Della Vedova, and Paola Bonizzoni. “PintronNext: a fast method for detecting the gene structure due to alternative splicing via ESTs, mRNAs, and RNA-Seq data”. In: *EURASNET Symposium on Regulation of Gene Expression through RNA Splicing*. 2012.
- [5] **Beretta, Stefano**, Paola Bonizzoni, Raffaella Rizzi, and Gianluca Della Vedova. “Identification of Alternative Splicing variants from RNA-seq Data”. In: *Next Generation Sequencing Workshop*. 2011.
- [6] **Beretta, Stefano**, Paola Bonizzoni, Raffaella Rizzi, and Gianluca Della Vedova. “Alternative Splicing from RNA-seq Data without the Genome”. In: *ISMB/ECCB - 8th Special Interest Group meeting on Alternative Splicing (AS-SIG)*. 2011.

Tesi di Dottorato di Ricerca

- [7] **Beretta, Stefano**. “Algorithms for next generation sequencing data analysis”. HDL: 10281/42355. PhD thesis. Università degli Studi di Milano-Bicocca, Feb. 2013.

EVENTI E CONVEGNI

Partecipazione a Comitati Organizzativi

- Comitato organizzativo della conferenza *Computability in Europe (CiE) - The Nature of Computation*, 1-5 Luglio 2013, Università degli Studi di Milano-Bicocca.

Relazione a Convegni Internazionali

- Partecipazione a *Bioinformatics and Biomedicine (BIBM)*, *2012 IEEE International Conference*, 4-7 Ottobre 2012, Philadelphia PA, USA, in qualità di relatore del lavoro [2].
- Partecipazione a *Next Generation Sequencing Workshop*, 12-14 Ottobre 2011, Bari, Italia, in qualità di relatore del lavoro [5].
- Partecipazione a *ISMB/ECCB - 8th Special Interest Group meeting on Alternative Splicing (AS-SIG)*, 15-16 Luglio 2011, Vienna, Austria, in qualità di relatore del lavoro [6].

Partecipazione a Convegni Internazionali

- Partecipazione a *RECOMB - Conference on Research in Computational Molecular Biology*, Aprile 2012, Barcellona, Spagna.

Scuole

- Frequenza della scuola estiva Lipari International Summer School on Bioinformatics and Computational Biology, dal titolo “*Biological Sequence Analysis and High Throughput Technologies*”, organizzata dall’Università degli Studi di Catania, 2011.

Visite di Ricerca

- *Visiting Scholar* presso la Technical University of Catalonia (UPC) in Barcellona, Spagna, da Aprile 2012 a Luglio 2012, sotto la supervisione di Prof. Gabriel Valiente.

ATTIVITÀ DIDATTICHE

Esercitazioni

- Insegnamento “Laboratorio di Algoritmi e Strutture Dati”, A.A. 2010/11.
Esercitazioni di laboratorio, 24 ore. Corso di Laurea in Informatica, Facoltà di Scienze MM. FF. NN., Univ. degli Studi di Milano-Bicocca.
- Insegnamento “Laboratorio di Informatica I”, A.A. 2010/11.
Esercitazioni di laboratorio, 24 ore. Corso di Laurea in Fisica, Facoltà di Scienze MM. FF. NN., Univ. degli Studi di Milano-Bicocca.
- Insegnamento “Laboratorio di Matematica e Informatica I”, A.A. 2010/11.
Esercitazioni di laboratorio, 24 ore. Corso di Laurea in Matematica, Facoltà di Scienze MM. FF. NN., Univ. degli Studi di Milano-Bicocca.
- Insegnamento “Laboratorio di Algoritmi e Strutture Dati”, A.A. 2009/10.
Esercitazioni di laboratorio, 12 ore. Corso di Laurea in Informatica, Facoltà di Scienze MM. FF. NN., Univ. degli Studi di Milano-Bicocca.

Docente di lezioni frontali

- Insegnamento “Informatica di Base 1”, A.A. 2011/12.
Lezioni frontali, 60 ore. Corso di Laurea in Lingue e Letterature Straniere, Facoltà di Lingue e Letterature Straniere, Univ. degli Studi di Bergamo.
- Insegnamento “Informatica di Base 1”, A.A. 2010/11.
Lezioni frontali, 60 ore. Corso di Laurea in Lingue e Letterature Straniere, Facoltà di Lingue e Letterature Straniere, Univ. degli Studi di Bergamo.

Altre esperienze didattiche di livello universitario

- Co-docenza nell’ambito dell’iniziativa “Learning Week”, Univ. degli Studi di Milano-Bicocca, 8-12 febbraio 2010.

SUPERVISIONE DI STAGE E TESI

Stage

- Co-supervisione di Stage, effettuato a conclusione del Corso di Laurea in Informatica, dal titolo “*Sviluppo di software per l’identificazione di variazioni strutturali da dati NGS*” svolto da Massimiliano Gilardi presso l’Univ. degli Studi di Milano-Bicocca, A.A. 2010/11.